基于图优化的视觉 SLAM 研究进展与应用分析 *

王录涛，吴林峰

(成都信息工程大学 计算机学院，成都 610225)

摘　要：同步定位与构图技术(simultaneous localization and mapping，SLAM)指机器人在构建环境地图的同时对自己的运动状态进行估计，是实现未知环境下机器人全自主运动的核心。为了对 SLAM 技术有更为全面的把握，在回顾过去三十年里视觉 SLAM 技术发展历程基础上，详细分析了视觉 SLAM 问题的本质与求解的复杂性。重点对在提高位姿估计精度、构建全局一致地图与提升算法求解效率上的最新研究成果进行了介绍，并对当前代表性的算法实现方案进行了分析与比较。针对未来大尺度环境、全生命周期应用需求，对现有算法框架的不足与最新研究趋势进行了归纳总结。最后，探讨了深度学习技术与视觉 SLAM 问题求解的关联性。

关键词：同步定位与构图；图优化；数据关联；稀疏化；深度学习

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.08.0509

Application analyses and research progress of graph-based visual SLAM

Wang Lutao, Wu Linfeng

(*School of Computer Science*, *Chengdu University of Information Technology*, *Chengdu* 610225, *China*)

Abstract: Simultaneous localization and mapping consists in the concurrent construction of a model of environment, and the estimation of the robot moving within it. It is one of the key problems for the robot to be completely autonomous in unknown environments. In order to give a comprehensive understanding of SLAM, we first give an overview of the progress of visual SLAM community has made over the last 30 years in this survey. We then present the non-linearity of the mathematical model and computational complexity in visual SLAM algorithms. We mostly focus on the latest achievements and approaches improving the accuracy of pose estimation, building a globally consistency representation of the environment and promoting computation efficiency. We also survey the failure modes of current visual SLAM algorithms for large scale, full lifecycle implementation and the different ways to address that. Finally, we discuss the potential connections between deep learning architectures and visual SLAM state estimations.

Key words: simultaneous localization and mapping; graph optimization; data association; sparsification; deep learning

0　引言

为实现未知环境下全自主运动，移动机器人必须解决环境感知与自身定位两个基本的问题。同步定位与地图构建技术(Simultaneous localization and mapping，SLAM)将机器人定位与地图构建融为一体，使机器人在缺乏环境先验信息的情况下，能依据自身搭载的传感器在运动过程中增量获取未知环境的特征信息，同时实现自我运动轨迹的精确估计[1]。相对激光 SLAM 技术，视觉 SLAM 系统硬件成本低、能获取环境纹理与色彩信息，因而具有更为广阔的应用前景。

迄今为止，视觉 SLAM 技术的发展过程可归纳为两个阶段。第一阶段从 1986 年到 2004 年，采用贝叶斯滤波技术求解 SLAM 问题。Smith 等[2]进行了开创性的研究，首先将 SLAM 问题系统地表示为随机估计问题。机器人位姿和地标被看做服从某种分布的随机变量，利用运动数据和观测数据，采用滤波理论实现系统状态预测与观测更新，同时实现地图的在线更新，具有较好的实时性。代表性算法有扩展卡尔曼滤波(Extended Kalman Filters, EKF)[3]、扩展信息滤波[4]、UKF[5]、Rao-Balckwellized 粒子滤波[6]、FastSlam[7]等。滤波方法假设状态估计的 Markov 性，仅用相邻帧信息对机器人状态进行估计，难以处理当前帧与历史帧的数据关联问题。SLAM 系统运动方程和观测方程是非线性函数，滤波方法采用一阶泰勒近似计算状态的后验概率，当系统存在强烈的非线性时，大的线性化估计误差将不可避免。此外，系统参数和观测存在的不确定性，也会造成误差的累积，

收稿日期：2018-08-06；修回日期：2018-10-09　　基金项目：四川省科技厅重点研发项目（18ZDYF3214）；成都市科技惠民技术研发项目（2016-HM01-00406-SF）

作者简介：王录涛（1979-），男，讲师，博士（后），主要研究方向为机器视觉、机器人同步定位与同步构图技术（wltuestc@163.com）；吴林峰（1991-），男，硕士研究生，主要研究方向为机器人定位与同步构图．

导致构建地图的不一致性。在算法实现上，滤波方法需存储、维护和更新状态变量的均值和方差，且存储容量与待估计状态变量呈平方关系。因此，滤波方法仅适用于计算资源受限，或估计变量较少的情况下。Paz 等人采用子地图分割的方法解决 EKF 的工作环境尺度问题，但仅能在 100 米的范围内有效工作[8]。

第二阶段(2004-2016)致力于 SLAM 理论分析与实现细节研究，关于 SLAM 问题的可观测性、收敛性与一致性的研究大量见诸于报道。降低对计算资源的需求，实现系统实时性工作的同时扩展应用环境尺度成为 SLAM 应用的关键[9]，而利用图优化技术，构建针对观测方程的非线性最小二乘目标函数，将机器人位姿与路标点作为待优化变量，采用牛顿法、Levenberg-Marquardt 算法迭代估计最优解，提高局部定位与路标点的估计精度，成为 SLAM 研究的主流。同时，SLAM 问题的稀疏性也逐渐被认知，使得全局一致解的获取成为可能。目前，在 50 km 的工作环境尺度下，具有 1%平移误差与 0.003 $deg \cdot m^{-1}$ 旋转误差的 SLAM 系统已经出现[10]。

经过三十多年的发展，视觉 SLAM 理论框架趋于稳定，算法定位精度、运动轨迹估计漂移、全局一致性环境地图构建等关键性问题在计算资源、运动模型、静态环境与性能等约束条件得以满足的情况下，得到了有效解决，但针对全生命周期、大尺度复杂环境下的高性能、低失效率应用，以及计算资源的自适应应用与构建基于任务驱动的感知模型等方面还缺乏更为有效的解决方案。Liu 等[11]对近期代表性的单目视觉 SLAM 技术进行了比较分析。Cadena [12]总结了 SLAM 的状态估计存在的问题、地图表示方法与研究进展，以及 SLAM 性能的理论保证，并对相关前沿课题进行了详细论述。Cuillaume [13]对 SLAM 在自动驾驶应用中存在的问题与解决途径进行了分析。

本文侧重对视觉 SLAM 技术的理论基础与全生命周期、大尺度环境工作存在的问题及解决途径进行深入剖析，首先给出了 SLAM 系统处理架构，结合理论基础分析图优化估计方法存在的问题，并对最新研究进展进行了阐述，然后归纳总结了现阶段的主流算法及可能达到的性能指标。最后，针对未来 SLAM 系统全生命周期、开放大尺度工作场景下的应用需求，探讨了存在的问题及解决途径，并对研究趋势进行了展望。

1　SLAM 数学模型

SLAM 技术可表述为在未知环境中，移动机器人利用搭载传感器获得的环境特征信息增量式构建环境地图，同时实现自身位姿的精确估计，如图 1 所示。

在图 1 中，机器人 N 个时刻的位姿 $x_1,\cdots,x_n$ 构成自身运动轨迹，$y_1,\cdots,y_l$ 为路标。根据 k 时刻的位姿 x_k 与在 x_k 对路标 y_j 的观测产生的观测数据 $z_{k,j}$，SLAM 问题模型可用一个运动模型和观测模型表示

$$\begin{cases} x_k = f(x_{k-1}, u_k, w_k) \\ z_{k,j} = h(y_j, x_k, \varepsilon_{k,j}) \end{cases} \tag{1}$$

其中: u_k 为运动传感器输入，w_k 为状态噪声，ε_k 为观测噪声。

根据 SLAM 数学模型，可将 SLAM 系统分为前端与后端两部分，如图 2 所示。前端一方面根据传感器数据提取环境特征，实现观测与路标、以及位姿的数据关联，另一方面为后端的非线性优化估计提供可靠初始值。数据关联是前端处理的关键，既要解决连续测量帧的特征关联，实现运动跟踪，也要负责当前测量与历史测量的关联，完成定位与构图功能。后端则在前端基础上进行位姿与路标点的推理与优化，实现全局一致性地图构建，改善定位精度估计。

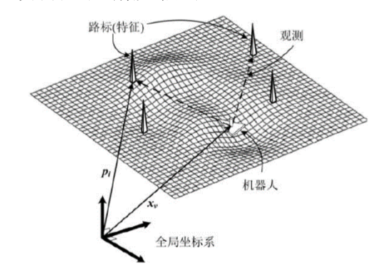


图 1　SLAM 问题示意图

Fig.1　Illustration of SLAM problem

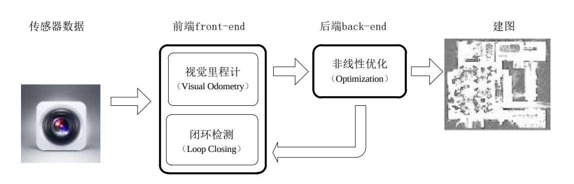
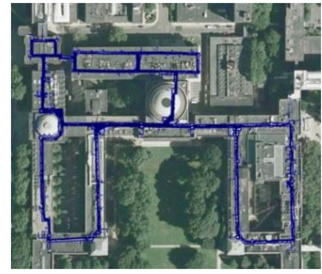


图 2 SLAM 系统框架

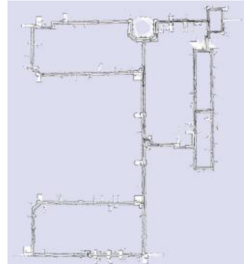
Fig.2　Architecture of SLAM system

图 3 给出了后端优化处理前后运动轨迹估计对比，采用的数据集为 MIT Killian Court。图 3(a)为 MIT Killian Court 光学照片与机器人运动路径标注。由于传感器测量误差与机器人位姿估计累积误差，机器人运动轨迹估计结果与实际标准值存在较大偏差，而后端优化则可以对估计误差进行修正，使运动轨迹估计与实际运动轨迹更为相符。



(a) MIT Killian Court 光学照片与机器人运动轨迹标注

(a) Optical map of MIT Killian Court with trajectory annotation

(b) 未经后端优化处理　(c)后端优化处理后

(b)without optimization　(c)with optimization

图 3　后端优化处理前后运动轨迹估计对比

Fig.3　Estimated trajectory with or without back-end optimization

2　基于图优化的 SLAM 问题求解

2.1　SLAM 问题描述

在有噪声的观测数据中精确估计状态变量是 SLAM 的基本问题。图优化技术将机器人轨迹与路标联合 $x=[x_1,\cdots,x_n,y_1,\cdots,y_l]$ 作为待估计变量，在给定观测数据 $z=\{z_k,k=1,\cdots,m\}$ 的情况下估计 x，使得后验概率分布 $\mathrm{P}(x\mid z)$ 最大。根据贝叶斯法则，求解最大后验概率等于最大化似然 $\mathrm{P}(z\mid x)$ 和先验概率 $\mathrm{P}(x)$ 的乘积

$$x_{MAP}{}^{*}=\arg\max_{x}\mathrm{P}(x\mid z)=\arg\max_{x}\mathrm{P}(z\mid x)\mathrm{P}(x) \tag{2}$$

在没有先验信息的情况下，$\mathrm{P}(x)$ 为常数，$x_{MAP}{}^{*}$ 简化为最大似然概率估计(Maximum Likelihood Estimation, MLE)

$$x^{*}=\arg\max_{x}\mathrm{P}(z\mid x) \tag{3}$$

在线性高斯假设下，图优化方法与 KF 方法估计结果一致，但对于非线性非高斯系统，图优化技术将位姿与路标变量作为一个整体，使用所有时刻采集数据进行联合估计，克服了 EKF 的线性化误差与噪声高斯分布假设，成为当今研究的主流。在计算资源允许的条件下，图优化方法也成为优选方法。

假设观测数据 Z 是独立的，最大后验概率的估计可分解为各次观测最大似然的乘积

$$\begin{aligned}x_{MAP}{}^{*}&=\arg\max_{x}\mathrm{P}(x)\prod_{k=1}^{m}\mathrm{P}(z_k\mid x)\\&=\arg\max_{x}\mathrm{P}(x)\prod_{k=1}^{m}\mathrm{P}(z_k\mid x_k)\end{aligned} \tag{4}$$

其中：$x_k\subseteq x$。假设传感器受高斯白噪声影响，观测值服从高斯分布，协方差矩阵为 Σ_k，则最大测量似然可表示为

$$\mathrm{P}(z_k\mid x_k)\propto\exp\left(-\frac{1}{2}\|h_k(x_k)-z_k\|_{\Sigma_k^{-1}}^{2}\right) \tag{5}$$

将式(5)代入式(4)，根据最大化后验概率等于最小化负对数似然函数，最大后验概率的估计可进一步显示的表述为求解机器人位姿与路标的联合估计，使得估计结果与观测的误差平方函数最小

$$\begin{aligned}x^{*}&=\arg\max_{x}-\log\left(\mathrm{P}(x)\prod_{k=1}^{m}\mathrm{P}(z_k\mid x_k)\right)\\&=\arg\min_{x}\sum_{k=0}^{m}\frac{1}{2}\|h_k(x_k)-z_k\|_{\Omega_k}^{2}\end{aligned} \tag{6}$$

其中：$h_k(\bullet)$ 是抽象的非线性函数，可用以表示惯性传感器、编码器、GPS、相机等数学模型。当噪声不满足标准正态分布时，式(6)中的度量误差 l_2 可用 l_1 等其他范数取代，为增加系统的稳健型，减少对外点的敏感度，也可用 Huber、Tukey 等损失函数代替 l_2 范数。

为方便表述，令 $e_k(x)\equiv h_k(x_k)-z_k$，$e=\left[e_1^{\mathrm{T}},\cdots,e_m^{\mathrm{T}}\right]^{\mathrm{T}}$ 表示 m 维误差向量，$\Omega\equiv diag\left(\Omega_1^{\mathrm{T}},\cdots,\Omega_m^{\mathrm{T}}\right)^{\mathrm{T}}$ 为误差权重矩阵，$\Omega=\Sigma_k^{-1}$，式(6)可用矩阵的形式表示为

$$x^{*}=\arg\min_{x}e^{\mathrm{T}}\Omega e \tag{7}$$

对于式(6)所表示的非线性最小二乘优化问题，一般可采用迭代线性化方式求解。在给定初值 x_0 时，对误差函数 $e(x)$ 在 x_0 附近进行一阶泰勒展开，求关于 x_0 增量 δx 的导数并使其等于零，得到增量方程

$$H\delta x=g \tag{8}$$

其中，$g=-\mathrm{J}(x)^{\mathrm{T}}e(x)$ 为系统信息向量，$H=\mathrm{J}(x)^{\mathrm{T}}\mathrm{J}(x)$ 为 Hessian 矩阵，$\mathrm{J}(x)=\partial e/\partial x$ 为雅可比矩阵。进而，可求得系统的解为：

$$x^{*}=x_0+\delta x \tag{9}$$

Gauss-Newton 法对上述过程进行迭代求解直到收敛，可获得参量的最优估计。为避免 δx 过大导致的近似误差增加，Levenberg-Marquardt 算法通过对 δx 添加置信域，引入松弛因子 λ 有效改善了 δx 求解的稳定性：

$$(H+\lambda I)\delta x=g \tag{10}$$

2.2　算法本质分析

SLAM 问题中的位姿估计由平移和旋转估计两部分构成。其中，旋转计算的非线性决定了 SLAM 本质是非线性优化问题。此外，非高斯噪声、变量维数巨大，以及静态场景假设往往得不到满足等多种因素的存在，SLAM 优化问题的目标函数(7)异常复杂，位姿与路标的估计极易陷入局部极小，而位姿与路标的错误估计将使得机器人定位产生漂移，无法获得全局一致的运动轨迹估计与地图构建，从而难以满足导航、环境重建等应用需要。图 4 给出了球体和圆环全局优化估计与估计陷入局部极小时的仿真对比。

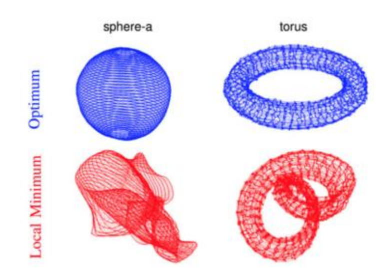


图 4　全局优化估计与局部优化估计仿真对比

Fig.4　Trajectory estimates resulting from global optimum and convergence to local minima

文献[14]利用视觉里程计与 3D 激光里程计的组合获取用于优化估计的初始值，在 KITTI 数据集上测试，平移计算精度达到 0.68%，旋转误差达到 0.0016 $\mathrm{deg}\cdot m^{-1}$。通过对观测特征点的筛选与跟踪，应用多帧图像特征点绝对差异求和、归一化互相关技术，抑制外点对优化估计结果的影响，单一视觉 SLAM 在 KITTI 数据集上的平移计算精度可达 1%，旋转误差低于 0.003 $\mathrm{deg}\cdot m^{-1}$ [15-17]，仍难以满足汽车自动导航等大尺度环境应用需求。

迭代求解的收敛性问题引发了人们从理论层次上对 SLAM 问题的深入研究，从而推动了算法研究的进展。Huang 等人最

早对 SLAM 问题的非凸性进行了研究，并对小尺度位姿图优化进行了探讨[18-20]。Knuth 等人研究了无环路闭合情况下状态估计误差的增长[21]。Carlone 等人对迭代求解的收敛性问题进行了持续研究，首先给出了 Gauss-Newton 法的收敛域的估计方法，并证明了 2D 场景下旋转估计的封闭性与唯一性，然后又提出根据 SLAM 问题的强对偶性，最大似然估计是唯一的，采用半正定规划(semidefinite programming, SDP)可得到位姿图的全局最优解[22-25]。Liu 等人采用凸松弛法避免求解陷入局部极小[26]。根据 SLAM 问题非线性主要来源于旋转计算的非线性，Carlone 等人先进行旋转估计，利用估计结果加速非线性迭代求解效果[27]。近期，为制定有效的安全紧要场景下的系统故障检测与失效恢复策略，Carlone 等人利用 SLAM 的 Lagrangian 对偶性，针对位姿估计结果的评估与验证，提出了有效的解决方案[28]。

2.3 SLAM 问题的求解

SLAM 待估计变量包含了所有机器人位姿与观测路标点。对于视觉 SLAM，路标点的数目远大于位姿数，δx 维数达到数千维，使得增量方程规模巨大。根据传感器的视野有限，可观测路标独立且有限这一事实，SLAM 问题具有良好的稀疏性，并可以用因子图直观表示，如图 5 所示。

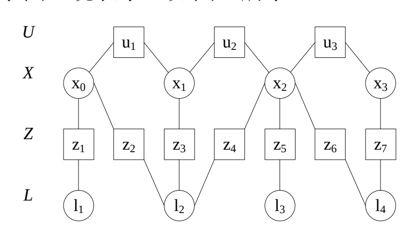


图 5　SLAM 因子图示意图

Fig.5　SLAM as a factory graph

因子图是一种二维无向图，用多个因子的概率分布的乘积估计联合概率分布[29]。SLAM 因子图由若干个节点与连接节点的边构成，圆形节点表示待优化状态变量，对应于机器人及系统在不同时刻的状态，矩形节点表示对路标的观测与控制输入，边描述了相邻位姿变换关系或对路标的观测误差。由于 i 时刻位姿对 j 路标的观测产生的误差项与其他时刻位姿与路标无关，雅可比矩阵 J(x) 中仅 $\partial e_{i,j}/\partial x_i$ 与 $\partial e_{i,j}/\partial y_j$ 项不为零：

$$J_{ij} = \left(0 \cdots 0 \quad \frac{\partial e_{i,j}}{\partial x_i} \quad 0 \cdots 0 \quad \frac{\partial e_{i,j}}{\partial y_j} \quad 0 \cdots 0 \right) \tag{11}$$

与图 4 对应的雅可比矩阵为

$$J = \begin{bmatrix} J_1 \\ \vdots \\ J_{10} \end{bmatrix} = \begin{bmatrix} J_{10} & J_{11} & & & & & & & \\ & J_{21} & J_{22} & & & & & & \\ & & J_{32} & J_{33} & & & & & \\ J_{40} & & & & J_{44} & & & & \\ J_{50} & & & & & J_{55} & & & \\ & J_{61} & & & & J_{65} & & & \\ & & J_{72} & & & J_{75} & & & \\ & & J_{82} & & & & J_{86} & & \\ & & J_{92} & & & & & J_{97} & \\ & & & J_{10,3} & & & & J_{10,7} & \end{bmatrix} \tag{12}$$

J(x) 对 H 的贡献为 $J_{i,j}^{\mathrm{T}} J_{i,j}$，$H$ 可表示为

$$H = \sum_{i,j} J_{i,j}^{\mathrm{T}} J_{i,j} \tag{13}$$

对 H 进行分块表示

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{14}$$

其中，H_{11}、H_{22} 分别与位姿和路标有关，$H_{21} = H_{12}^{\mathrm{T}}$。由于路标数远远大于位姿数，采用 Schur 消元法消去 H_{12}，可得关于位姿的增量方程

$$[H_{11} - H_{12} H_{22}^{-1} H_{12}^{\mathrm{T}}] \delta x_c = x_c - H_{12} H_{22}^{-1} x_p \tag{15}$$

其中，$x_c = [x_1, \cdots, x_n]$，$x_p = [y_1, \cdots, y_l]$。路标的增量方程为

$$\delta x_p = H_{22}^{-1}(x_p - H_{12}^{\mathrm{T}} \delta x_c) \tag{16}$$

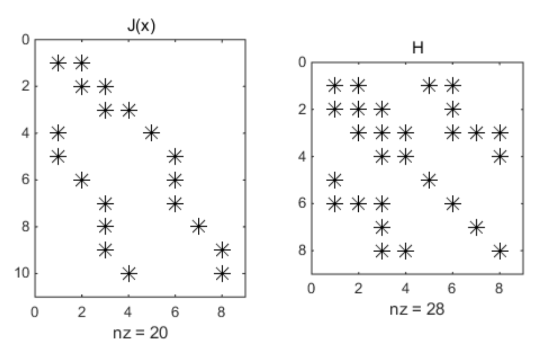


图 6　雅可比矩阵与 Hessian 矩阵的稀疏性

Fig.6　Sparsity of the Jacobian and Hessian matrix

图 6 给出了图 5 所示的由 4 位姿与 4 路标构成的 SLAM 问题的雅可比矩阵 J(x) 与 Hessian 矩阵 H 的稀疏性表示。由于 H 的稀疏性，且路标矩阵 H_{22} 为对角阵，位姿增量方程求解中，H_{22}^{-1} 易于求解，同时由于在实际问题求解过程中，位姿数远小于路标数 l，δx_c 求解规模相对于原问题大大简化。

利用 Hessian 矩阵的稀疏性，采用 Schur 消元与线性分解，然后计算增量方程，显著提高了 SLAM 问题优化求解的效率。目前，已涌现出众多因子图求解实现框架，如 g2o[30]、TORO[31]、HOG-Man[32]、COP-SLAM[33]等，在普通 PC 处理器已可实现数千维变量的实时求解，大大促进了 SLAM 技术从研究走向应用的进程。

3　图优化 SLAM 算法实现框架

图优化技术为 SLAM 系统的性能提供了理论保证，但 SLAM 系统的实现还依赖路标信息感知、特征数据关联、地图表示等关键技术，并与程序的设计等工程应用问题紧密相连，本节对现有代表性的基于图优化的 SLAM 系统实现方案与性能指标进行分析。

1) PTAM

PTAM 的提出是视觉 SLAM 发展过程中的具有革新意义的事件。PTAM[34]确立了视觉 SLAM 的基本实现框架，提出了两

大创举：1)首次采用非线性优化技术替代传统滤波技术作为 SLAM 后端的实现方案；2)引入关键帧机制，使对地图的优化可以整合到实时计算中。在方案实现时，设计 2 个独立的线程分别实现相机位姿的跟踪与地图的构建。位姿跟踪线程实时响应图像数据，地图构建线程专注于地图的建立、维护和更新。由于仅需维护视频关键帧与关键帧稳定观测的地标点，使得目标函数的优化得以高效求解。但是，当地图构建或优化过慢易导致跟踪线程产生跟踪丢失。Pire 等采用立体相机对 PTAM 技术进行改进，将立体约束用于路标特征的初始化、跟踪与构图，在系统中增加了实时环路检测与修正模块，并采用局部并行光束平差技术(BA)优化局部地图，实现了实时求解，同时位姿估计精度也得到了提升[35]。

2) ORB-SLAM

ORB-SLAM 继承了 PTAM 的后端非线性优化实现方案与关键帧处理机制，支持单目、双目、RGB-D 三种视频输入模式，是现代 SLAM 系统中最为完善、易用的系统之一[36,37]。采用 ORB 特征实现目标特征匹配与跟踪，在普通 CPU 上实现实时计算的同时又兼具良好的旋转与缩放不变性。ORB 特征描述子的实时提取与离线构建的 ORB 字典的运用，使大尺度运动时的回环检测与目标重定位成为可能。ORB-SLAM 代码实现采用三线程：实时特征跟踪线程根据关键帧匹配路标 ORB 特征描述子，粗略计算路标位置与相机位姿；局部 BA 线程维护局部共视图，求解精细的路标位置与相机位姿；全局环路检测与优化线程消除状态估计累积误差，获取全局一致运动轨迹估计。相比 ORB-SLAM，ORB-SLAM2 增加了地图重用功能，解决地图构建模块失效时的定位漂移问题，定位精度与大尺度范围工作时的稳健型也得以提高[38]。

3) LSD-SLAM：

LSD-SLAM 采用直接法，根据像素的梯度信息估计相机运动与构建半稠密地图，避免了关键点的提取与描述子的计算[39]。利用极线上等距离 5 个点的 SSD 度量保证跟踪的稳定性；在深度估计时，采用随机数初始化深度，在估计完成后进行深度均值归一化，调整尺度；在度量深度不确定性时，将几何关系与极线、深度夹角关系归纳成光度不确定相；在后端优化中考虑不同场景尺度，减小尺度漂移。直接法跟踪对特征缺失区域不敏感，但对相机内参和曝光非常敏感，在相机快速移动时容易丢失，此外，在环路检测时，仍依赖于特征点的计算。S-LSD-SLAM 采用双目立体相机估计场景深度，解决单目相机的尺度漂移问题，利用仿射光照校正技术使得在相邻关键帧光照存在强烈差异时，光度残差近似不变[40]。

图 7 为特征地图与半稠密重建对照。半稠密地图根据灰度图中的梯度差异建模，将物体的边缘或表面纹理部分在地图中进行显示，比稀疏地图具有更多的信息。

4) SVO

SVO 综合使用直接法与特征法计算路标特征与实现目标跟踪。SVO 用特征点周围 4x4 图像块估计相机位姿与路标位置，既不提取描述子，也不处理稠密或半稠密信息，降低了对 CPU 计算能力的要求，可实现在无人机、手持 AR/VR 等设备上的实时应用[41,42]。SVO 的显著贡献是提出深度滤波技术并应用于路标点的位置估计。但是，由于不包含后端优化和环路检测，且不存在建图功能，SVO 位姿估计存在累积误差，在跟踪丢失后难以实现重定位。

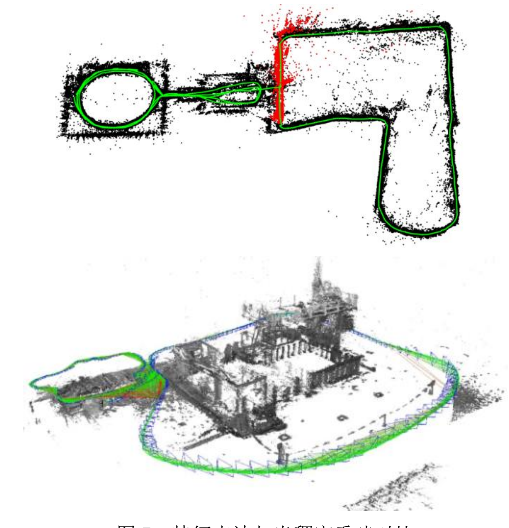


图 7　特征点法与半稠密重建对比

Fig.7　Feature-based map vs. semi-dense reconstruction

5) SOFT

SOFT 采用双目立体相机，通过对稳定跟踪特征的子集的精细选择减少定位漂移[43]。SOFT 在当前帧的一个小的窗口内提取 Blob 特征与角点特征，通过非极大值抑制技术判断特征点间的对应关系，并采用归一化互相关去除外点，确定稳定匹配特征。将运动估计分为旋转与平移两个部分，旋转估计采用 5 点法与 RANSAC[44]，平移估计采用最小重投影误差技术，减少匹配误差。此外，还可采用 IMU 信息抑制外点与优化旋转估计，可在 ARM 平台上实现高精度位姿的实时估计。

上述 5 种算法的位姿估计精度与实时性对比如表 1 所示。

表 1　SLAM 实现方案性能对比

Table 1　Performance comparison of SLAM algorithms

实现方案	位姿估计精度		每帧运行时间(s)	运行环境
	平移	旋转(deg/m)		
S-PTAM	1.35%	0.0023	0.08	4 cores, 2.2GHz, C/C++
ORB-SLAM2	1.15%	0.0027	0.06	2 cores, 3.5GHz, C/C++
S-LSD-SLAM	1.20%	0.0033	0.07	1 cores, 3.5GHz, C/C++
SVO2	0.94%	0.0021	0.2	1 cores, 2.5GHz, C/C++
SOFT	0.88%	0.0022	0.1	2 cores, 2.5GHz, C/C++

4　应用及展望

1）算法稳健型

算法稳健性是视觉 SLAM 系统全生命周期工作需解决的首要问题。算法设计的局限性与硬件相关问题，如传感器失效、驱动器误差等，使得当前 SLAM 系统只能在特定环境下、特定硬件平台上达到特定的性能要求[45-48]。

在众多影响视觉 SLAM 系统稳定工作的因素中，数据关联的计算涉及到传感器测量之间、传感器测量与地图特征之间或地图特征之间的对应关系，直接决定了 SLAM 问题求解的准确性与实时性。对于错误数据关联导致的算法失效可以从 SLAM 系统前端与后端两个方面入手加以解决。在前端，满足传感器采样速率远高于机器人运动姿态变化的情况下，特征描述子法与光流法[49]均可以有效的实现关键帧观测路标的准确跟踪。为降低累积估计误差，需将当前观测数据与历史观测数据实时关联，实现环路检测。词袋法[50]是目前最为有效的实现环路检测技术之一，可显著增强变量估计的一致性。其基本思想是从图像中提取的局部特征进行聚类，将连续变化特征量化为离散的单词，然后采用词的统计直方图对场景进行描述。在算法实现上则通过应用分层词汇树提高大规模数据集中的特征检索效率，实现大尺度环境实时环路检测[51,52]。但对于存在强烈光照变化的场景，视觉字的准确匹配难以实现[53]。将不同视觉外观纳入统一表征框架[54]，或者采用路标视觉外观信息与空间位置关系融合[55]，降低光照变化对场景识别的影响，增强环路检测的稳健性，是最近研究的热点问题。文献[56]对视觉场景识别方法进行了详细的归纳与总结。

尽管前端的环路检测技术取得了长足进步，但视觉混叠导致的环路检测错误对后端估计失效的影响仍难以避免，增强后端算法对虚假数据关联的抵御能力显得至关重要。常用的解决途径包括：1)利用先验证知识，检测错误的闭合环路，在进行优化计算之前抑制外点的影响；2)根据优化估计残差验证环路检测的正确性[57-61]。

基于图优化的 SLAM 问题的本质是求解非线性、非凸优化问题，算法求解器对于初始值非常敏感，且极易陷入局部极小点，导致估计结果存在巨大偏差。理想的 SLAM 系统应能实时评估算法估计结果，并具有从算法失效状态实现自我恢复的能力。紧密整合 SLAM 前端与后端，是提高 SLAM 算法稳健性的有效解决途径，但相关研究工作还有待进一步展开。此外，利用视觉、IMU、GNSS、3D 激光探测器构建定制地图，并实现在已有地图中的重定位技术，增强 SLAM 技术在特定场景的应用，如自动驾驶等，也是当前 SLAM 技术研究的热点问题之一[62]。

2）可扩展性

可扩展性是 SLAM 系统实现长时间、大尺度应用，如自动驾驶、海底环境检测、大规模精准农业等，需解决的另一个重要问题。随着工作时间的延长与探索范围的拓展，SLAM 位姿图规模将无限增加而线性迭代求解对存储资源的需求与待估计变量的数量成正比，导致系统不可实现[63]。降低因子图优化的复杂度，使得问题求解对计算与存储资源的需求保持恒定，主要途径是稀疏化方法与子图处理技术。

基于 Markov 毯理论，通过边与节点的边缘化可获得因子图的稀疏逼近。Lla 等采用信息论方法控制因子图中节点与边增添，仅增加非冗余节点与信息量含量高的边[64]。Jonhannsson 等通过对已有节点引入新的约束尽量避免新的节点的添加，使得因子图的扩展只与机器人探索区域有关，而与工作时间无关[65]。Kretzschmar 等针对位姿图优化问题，研究了基于信息论的节点与边的边缘化准则[66]。Carlevaris 与 Mazuran 等介绍了广义线性约束因子与相应的非线性图稀疏化方法[67,68]。实现因子图稀疏化的另一个途径是通过对连续时间运动轨迹的估计减少待估计参数，如利用三次样条曲线、B 样条曲线等，采用滑动窗口或批处理模式表示机器人运动路径[69-71]。Chi 等用高斯过程表示取代基本样条曲线表示法，稀疏化因子图中的节点是真实的机器人位姿，其他位姿可通过计算给定时间的后验均值的插值获得[72]。

子图处理技术将因子图分解为多个子图，利用多个处理器分布式计算局部因子图优化进而实现全局因子图的优化[73]。Ni[74]等提出基于二进制树结构的子图构建与组织方法，Grisett[75]对子图进行等级划分，新的观测仅更新最高等级与低等级受影响区域。将大尺度场景分解成多个小的区域，利用多个机器人分别进行构图为子图处理技术提供了另一种思路[76,77]。多机器人构图又可分为中心融合处理与分布式处理两类。中心融合处理采用中心处理单元对子图信息进行融合，而分布式处理通过机器人间相互通信保持建图信息的一致性。文献[80]对多机器人构图技术进行了详细的分析。

目前，SLAM 算法可扩展性研究仍集中于简化因子图优化的复杂度，对机器人大尺度环境工作涉及的其他众多问题，如语义地图构建、分布式建图的稳健性与通过构图实现环境的深度感知使机器人具有类人智能等的研究还有待进一步展开。

3）深度学习

深度学习技术的发展引起了计算机视觉领域的一场变革。目前，利用深度学习技术解决 SLAM 求解问题的研究已经展开。Costante 等采用表示学习方法取代视觉里程计中的几何约束，根据相邻两帧图像输入估计机器人位姿[78]。Eigen 等通过对路标点赋予语义信息，联合利用路标点的位置信息与语义信息提高数据关联的正确率，从而改进位姿与路标点位置的估计精度[79]。刘等利用深度学习网络，实现了基于单帧图像的场景深度估计[80,81]。此外，语义信息的引入为环路检测、地图表示等带来更多的条件[82,83]。

通过构建复杂的 CNN 网络，深度学习可实现目标特征的多层次提取，进而实现环境信息的深度感知，而传统计算机视觉技术对环境的认知仍停留在像素或特征点等浅层次特征提取上，因此，作为感知工具，深度学习技术可以解决传统计算机视觉技术无法解决的问题。但是对于 SLAM 技术，环境感知是为定位与构图服务的，是否能用深度学习技术实现端到端的 SLAM 系统仍处于探索之中。此外，利用场景先验信息可以显著提升 SLAM 系统的性能，但将深度学习网络输出的不确定性对 SLAM 视觉几何处理的影响的研究还有待展开。

未来的 SLAM 系统工作于开放的环境，需要对环境进行持续的探索，因此应具有全生命周期的学习能力，而深度学习的



成功应用依赖于对固定目标类的大量数据与标注样本的训练，因此，改进深度学习技术的在线学习与自适应能力，应是实现具有类人智能的 SLAM 系统的首要解决问题之一。此外，开发轻量级的网络，满足 SLAM 系统的嵌入式应用需求，也为深度学习与 SLAM 技术的结合带来了诸多挑战。

5 结束语

本文回顾了 SLAM 技术近三十年的发展历程，详细分析了基于图优化的视觉 SLAM 的理论模型与算法实现，探讨了问题求解中存在的问题并给出了最新的理论研究成果。目前，SLAM 理论研究已趋于完善，满足大尺度环境定位与构图的算法框架相继被提出，但针对全生命周期、大环境尺度工作应用需求，其算法的稳健性与可扩展性设计仍面临着诸多挑战。构建大尺度、全自主运行的 SLAM 系统，满足自动驾驶、海洋测绘、大规模精准农业等复杂场景应用是现代 SLAM 技术发展的目标，将传统 SLAM 技术与深度学习等新技术结合建立对工作环境的语义层次的深度感知，同时融合 GNSS、惯性测量、激光雷达等传感技术，增强系统复杂环境下工作的稳健性，则为该目标的实现提供了有效地解决途径。

参考文献：

[1] Durrant W H, Bailey T. Simultaneous localization and mapping: Part I [J]. IEEE Robotics Automat Mag, 2006, 13 (3): 108-117.

[2] Smith R C, Cheeseman P. On the estimation and representations of spatial uncertainty [J]. International Journal of Robotics Research, 1986, 5 (12): 56-68.

[3] 何俊学，李战明. 基于视觉的同时定位与地图构建方法综述 [J]. 计算机应用研究, 2010, 27 (8): 2839-2843. (He Junxue, Li Zhanming. Survey of vision-based approach to simultaneous localization and mapping [J]. Application Research of Computers, 2010, 27 (8): 2839-2843.)

[4] Thrun S, Liu Yufeng, Koller D, *et al.* Simultaneous localization and mapping with sparse extended information filters [J]. International Journal of Robotics Research, 2003, 23 (8): 693-716.

[5] Holmes S A, Klein G, Murray D W. An O (N^2) square root unscented kalman filter for visual simultaneous localization and mapping [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2009, 31 (7): 1251-1263.

[6] Doucet A, Freitas N D, Murphy K, *et al.* Rao-blackwellised particle filtering for dynamic Bayesian networks [C]// Proc of the 16th Conference on Uncertainty in Artificial Intelligence. 2000: 176-183.

[7] Montemerlo M, Thrun S, koller D, *et al.* Fast SLAM: a factored solution to the simultaneous localization and mapping problem [C]// Proc of AAAI National Conference on Artificial Intelligence. 2002: 593-598.

[8] Paz L M, Jensfelt P, Tardos J D, *et al.* EKF SLAM updates in O (n) with divide and conquer SLAM [C]// Proc of IEEE International Conference on Robotics and Automation. 2007: 1657-1663.

[9] Gamini D, Huang Shoudong, Wang Zhan, *et al.* A review of recent developments in simultaneous localization and mapping [C]// Proc ofIEEE International Conference on Industrial and Information Systems. 2011: 477-482.

[10] Persson M, Piccini T, Felsberg M, *et al.* Robust stereo visual odometry from monocular techniques [C]// Proc of Intelligent Vehicles Symposium. 2015: 686-691.

[11] Liu Haomin, Zhang Guofeng, Bao Hujun. A survey of monocular simultaneous localization and mapping [J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28 (6): 854-871.

[12] Cadena C, Carlone L, Carrillo H, *et al.* Past, present, and future of simultaneous localization and mapping: Toward the Robust-Perception Age [J]. IEEE Trans on Robotics, 2016, 32 (6): 1309-1332.

[13] Cuillaume B, Zayed A, Li Yu, *et al.* Simultaneous locationlization and mapping: a survey of current trends in autonomous driving [J]. IEEE Trans on Intelligent Vehicles, 2017, 2 (3): 194-230.

[14] Zhang Ji, Singh S. Visual-lidar odometry and mapping: low-drift, robust, and fast [C]// Proc of IEEE International Conference on Robotics and Automation. 2015: 2174-2181.

[15] Buczko M, Willert V. How to distinguish inliers from outliers in visual odometry for high-speed automotive applications [C]// Proc of Intelligent Vehicles Symposium. 2016: 478-483.

[16] Persson M, Piccini T, Felsberg M, *et al.* Robust stereo visual odometry from monocular techniques [C]// Proc of Intelligent Vehicles Symposium. 2015: 686-691.

[17] Cvišić I, Petrović I. Stereo odometry based on careful feature selection and tracking [C]// Proc of European Conference on Mobile Robots. 2015: 1-6.

[18] Huang Shoudong, Lai Yu, Frese U, *et al.* How far is SLAM from a linear least squares problem? [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2010: 3011-3016.

[19] Huang Shoudong, Wang Heng, Frese U, *et al.* On the number of local minima to the point feature based SLAM problem [C]// Proc of IEEE International Conference on Robotics and Automation. 2012: 2074-2079.

[20] Huang Shoudong, Dissanayake G. A critique of current developments in simultaneous localization and mapping [J]. International Journal of Robotics Research, 2016, 13 (5): 1-13.

[21] Knuth J, Barooah P. Error growth in position estimation from noisy relative pose measurements [J]. Robotics & Autonomous Systems, 2013, 61 (3): 229-244.

[22] 梁明杰，闵华清，罗荣华. 基于图优化的同时定位与地图创建综述 [J]. 机器人, 2013, 35 (4): 500-512. (Liang Mingjie, Min Huaqing, Luo Ronghua. Graph-based SLAM: A Survey [J]. Robot, 2013, 35 (4): 500-512.)

[23] Carlone L, Censi A. From angular manifolds to the integer lattice: guaranteed orientation estimation with application to pose graph optimization [J]. IEEE Trans on Robotics, 2014, 30 (2): 475-492.

[24] Carlone L, Rosen D M, Calafiore G, *et al.* Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions [C]// Proc of IEEE



International Conference on Intelligent Robots and Systems. 2015: 125-132.

[25] Carlone L, Calafiore G C, Tommolillo C, *et al.* Planar pose graph optimization: duality, optimal solutions, and verification [J]. IEEE Trans on Robotics, 2016, 32 (3): 545-565.

[26] Liu Mingjie, Huang Shoudong, Dissanayake G, *et al.* A convex optimization based approach for pose SLAM problems [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2012: 1898-1903.

[27] Carlone L, Aragues R, Castellanos J A, *et al.* A fast and accurate approximation for planar pose graph optimization [J]. International Journal of Robotics Research, 2014, 33 (7): 965-987.

[28] Carlone L, Tron R, Daniilidis K, *et al.* Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization [C]// Proc of IEEE International Conference on Robotics and Automation. 2015: 4597-4604.

[29] Kschischang F, Frey B, Loeliger H. A. Factor graphs and the sum-product algorithm [J]. IEEE Trans on Information Theory, 2001, 47 (2): 498-519.

[30] Kümmerle R, Grisetti G, Strasdat H, *et al.* G2o: A general framework for graph optimization [C]// Proc of IEEE International Conference on Robotics and Automation. 2011: 3607-3613.

[31] Grisetti G, Grzonka S, Stachniss C, *et al.* Efficient estimation of accurate maximum likelihood maps in 3D [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2007: 3472-3478.

[32] Grisetti G, Kummerle R, Stachniss C, *et al.* Hierarchical optimization on manifolds for online 2D and 3D mapping [C]// Proc of IEEE International Conference on Robotics and Automation. 2010: 273-278.

[33] Dubbelman G, Browning B. Closed-form Online Pose-chain SLAM [C]// Proc of IEEE International Conference on Robotics and Automation. 2013: 5190-5197.

[34] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces [C]// Proc ofIEEE and ACM International Symposium on Mixed and Augmented Reality. 2007: 1-10.

[35] Pire T, Fischer T, Civera J, *et al.* Stereo parallel tracking and mapping for robot localization [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2015: 1373-1378.

[36] Mur-Artal R, Montiel J M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. IEEE Trans on Robotics, 2015, 31 (5): 1147-1163.

[37] Mur-Artal R, Tardós J D. Fast relocalization and loop closing in keyframe-based SLAM [C]// Proc of IEEE International Conference on Robotics and Automation. IEEE Computer Society Press, 2014: 846-853.

[38] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source slam system for monocular, stereo, and RGB-D cameras [J]. IEEE Trans on Robotics, 2016, 33 (5): 1255-1262.

[39] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-Scale direct monocular SLAM [C]// Proc of the 13th European Conference on Computer Vision. 2014: 834-849.

[40] Engel J, Stückler J, Cremers D. Large-scale direct SLAM with stereo cameras [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2015: 1935-1942.

[41] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry [C]// Proc of IEEE International Conference on Robotics and Automation. 2014: 15-22.

[42] Cvišić I, Petrović I. Stereo odometry based on careful feature selection and tracking [C]// Proc of European Conference on Mobile Robots. 2015: 1-6.

[43] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Readings in Computer Vision, 1987, 24 (6): 726-740.

[44] 杨海燕，罗文超，刘国栋. 基于 SURF 算法和 SC-RANSAC 算法的图像配准 [J]. 计算机应用研究，30 (5): 1586-1588. (Yang Haiyan, Luo Wenchao, Liu Guodong. Image registration based on SURF algorithm and SC-RANSAC algorithm [J] , Application Research of Computers, 30 (5): 1586-1588.)

[45] Frese U. Interview: is SLAM solved? [J]. Künstliche Intelligenz, 2010, 24 (3): 255-257.

[46] Ackeerman E. Dyson's robot vacuum has 360-degree camera, tank treads, cyclone suction [EB/OL]. (2014-09-04) . https://spectrum. ieee. org/automaton/robotics/home-robots/dyson-the-360-eye-robot-vacuum.

[47] Jonathan H. New experimental features for Daydream. [EB/OL]. (2018-09-21) , https://developers. googleblog. com/2018/09/new-experimental-features-for-daydream. html.

[48] Kuka. KUKA automates laser system for automotive supplier Proseat [EB/OL]. (2017-09) . https://www. kuka. com/en-de/industries/solutions-database/2017/09/solution-systems-proseat.

[49] Scaramuzza D, Fraundorfer F. Visual Odometry: Part I: The first 30 years and fundamentals [J]. IEEE Robotics & Automation Magazine, 2011, 18 (4): 80-92.

[50] Galvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences [J]. IEEE Trans on Robotics, 2012, 28 (5): 1188-1197.

[51] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006: 2161-2168.

[52] Curless B, Levoy M. A volumetric method for building complex models from range images [C]// Proc of Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1996: 303-312.

[53] 曾文静，张铁栋，姜大鹏. SLAM 数据关联方法的比较分析 [J]. 系统工程与电子技术，2010, 32 (4): 860-864. (Zeng Wenjing, Zhang Tiedong, Jiang Daou. Analysis of data association methods of SLAM [J]. Systems Engineering and Electronics, 2010, 32 (4): 860-864.)

[54] Churchill W, Newman P. Experience-based navigation for long-term localisation [J]. International Journal of Robotics Research, 2013, 32 (14): 1645-1661.

[55] Ho K L, Newman P. Loop closure detection in SLAM by combining visual



and spatial appearance [J]. Robotics & Autonomous Systems, 2006, 54 (9): 740-749.

[56] Lowry S, Sünderhauf N, Newman P, *et al.* Visual place recognition: A Survey [J]. IEEE Trans on Robotics, 2016, 32 (1): 1-19.

[57] Sabatta D, Scaramuzza D, Siegwart R. Improved appearance-based matching in similar and dynamic environments using a Vocabulary tree [J]. 2010, 58 (8): 1008-1013.

[58] Sunderhauf N, Protzel P. Towards a robust back-end for pose graph SLAM [C]// Proc of IEEE International Conference on Robotics and Automation. 2012: 1254-1261.

[59] Latif Y, Cadena C, Neira J. Robust loop closing over time for pose graph SLAM [J]. International Journal of Robotics Research, 2013, 32 (14): 1611-1626.

[60] Olson E, Agarwal P. Inference on networks of mixtures for robust robot mapping [J]. International Journal of Robotics Research, 2013, 32 (7): 826-840.

[61] Carlone L, Censi A, Dellaert F. Selecting good measurements via ℓ 1, relaxation: A convex approach for robust estimation over graphs [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2014: 2667-2674.

[62] Wolcott R W, Eustice R M. Visual localization within LIDAR maps for automated urban driving [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2014: 176-183.

[63] Dellaert F, Carlson J, Ila V, *et al.* Subgraph-preconditioned conjugate gradients for large scale SLAM [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2010: 2566-2571.

[64] Ila V, Porta J M, Andrade-Cetto J. Information-based compact pose SLAM [J]. IEEE Trans on Robotics, 2010, 26 (1): 78-93.

[65] Johannsson H, Kaess M, Fallon M, *et al.* Temporally scalable visual SLAM using a reduced pose graph [C]// Proc of IEEE International Conference on Robotics and Automation. 2012: 54-61.

[66] Kretzschmar H, Stachniss C, Grisetti G. Efficient information-theoretic graph pruning for graph-based SLAM with laser range finders [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2011: 865-871.

[67] Carlevaris-Bianco N, Eustice R M. Eneric factor-based node marginalization and edge sparsification for pose-graph SLAM [C]// Proc of IEEE International Conference on Robotics and Automation. 2013: 5748-5755.

[68] Mazuran M, Burgard W, Tipaldi G D. Nonlinear factor recovery for long-term SLAM [J]. International Journal of Robotics Research, 2016, 35 (1): 50-72.

[69] Bibby C, Reid I. A hybrid SLAM representation for dynamic marine environments [C]// Proc of IEEE International Conference on Robotics and Automation. 2010: 257-264.

[70] Furgale P, Barfoot T D, Sibley G. Continuous-time batch estimation using temporal basis functions [C]// Proc of IEEE International Conference on Robotics and Automation. 2012: 2088-2095.

[71] Raguram R, Chum O, Pollefeys M, *et al.* USAC: a universal framework for random sample consensus [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2013, 35 (8): 2022-2038.

[72] Tong Chihay, Furgale P, Barfoot T D. Gaussian process gauss-newton for non-parametric simultaneous localization and mapping [J]. International Journal of Robotics Research, 2013, 32 (5): 507-525.

[73] Bosse M, Newman P M, Leonard J J, *et al.* Simultaneous localization and map building in large-scale cyclic environments using the atlas framework [J]. International Journal of Robotics Research, 2004, 23 (23): 1113-1139.

[74] Ni Kai, Dellaert F. Multi-level submap based SLAM using nested dissection [C]// Proc of IEEE International Conference on Intelligent Robots and Systems. 2010: 2558-2565.

[75] Grisetti G, Kummerle R, Stachniss C, *et al.* Hierarchical optimization on manifolds for online 2D and 3D mapping [C]// Proc of IEEE International Conference on Robotics and Automation. 2010: 273-278.

[76] Dong Jing, Nelson E, Indelman V, *et al.* Distributed real-time cooperative localization and mapping using an uncertainty-aware expectation maximization approach [C]// Proc of IEEE International Conference on Robotics and Automation. 2015: 5807-5814.

[77] Riazuelo L, Civera J, Montiel J M M. C 2 TAM: a cloud framework for cooperative tracking and mapping [J]. Robotics & Autonomous Systems, 2014, 62 (4): 401-413.

[78] Costante G, Mancini M, Valigi P, *et al.* Exploring representation learning with cnns for frame-to-frame ego-motion estimation [J]. IEEE Robotics & Automation Letters, 2015, 1 (1): 18-25.

[79] Eigen D, Fergus R. Predicting Depth, Surface normals and semantic labels with a common multi-scale convolutional architecture [C]// Proc of IEEE International Conference on Computer Vision. 2016: 2650-2658.

[80] Liu Fayao, Shen Chunha, Lin Guosheng. Deep convolutional neural fields for depth estimation from a single image [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015: 07-12.

[81] Cadena C, Dick A, Reid I D. Multi-modal auto-encoders as joint estimators for robotics scene understanding [C]// Robotics: Science and Systems. 2016: 78-83.

[82] Bao S Y, Bagra M, Chao Yuwei, *et al.* Semantic structure from motion with points, regions, and objects [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2703-2710.

[83] Dame A, Prisacariu V A, Ren Carl Y, *et al.* Dense reconstruction using 3D object shape priors [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2013: 1288-1295.